

# Use of Informatics to Understand Prostate Cancer Development and Progression

Scott A. Tomlins

July 19, 2006

Arul Chinnaiyan Laboratory

University of Michigan Medical School



# Background

- Prostate Cancer (PCA) is most common non-cutaneous malignancy in Western men
- PCA shows heterogeneity at the biological, clinical and histological level
  - The relationship of precursors to PCA is unclear
  - Genetic lesions driving PCA development are unknown
  - Clinical and pathological parameters identify tumors that are likely to recur
  - Clinically and pathologically identical tumors can have drastically different courses
- **Genes and processes controlling the development and progression of prostate cancer are unknown**

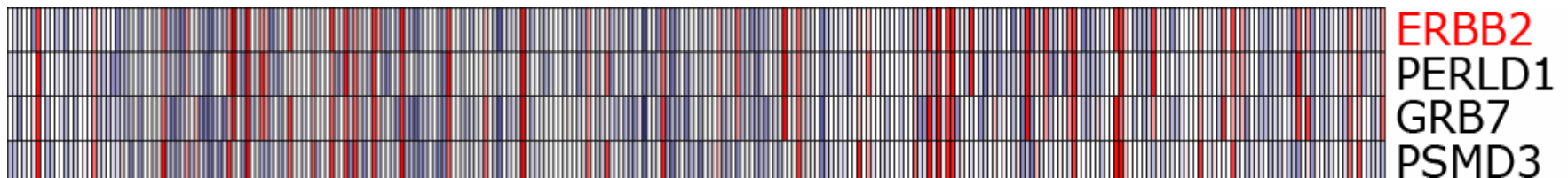


# A Bioinformatics Approach to Discover Candidate Oncogenes

- Gene expression profiling has become a common method to analyze the cancer transcriptome
- However, few causal cancer genes have been discovered using DNA microarrays
- We reasoned that chromosomal aberrations that result in marked over-expression of an oncogene should be evident in transcriptome data

# A Bioinformatics Approach to Discover Candidate Oncogenes

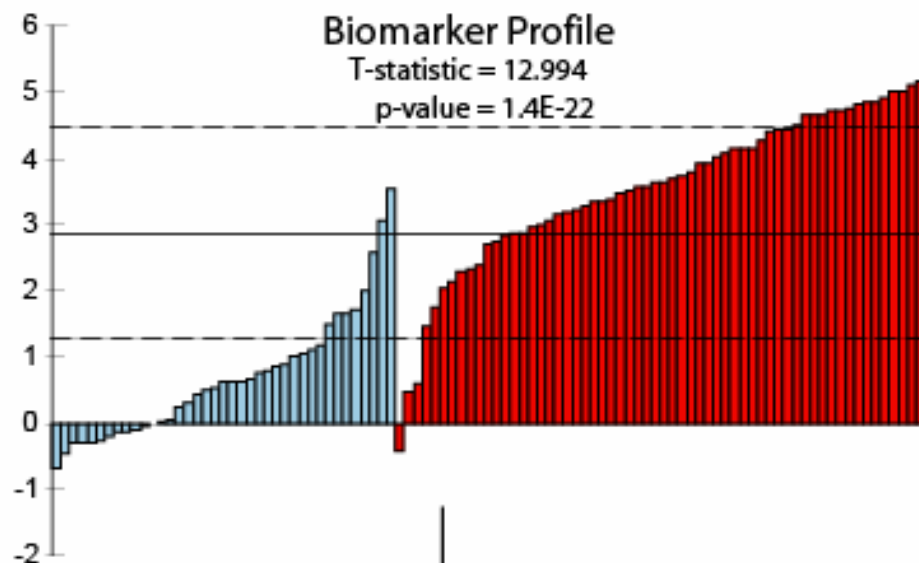
- Most analytical methods (t-test) divide samples into two classes “normal” vs. “cancer” and look for genes commonly over-expressed in class 2 vs. class 1.
- Oncogenic events are often heterogeneous
  - ERBB2/HER2 amplification in 20% of breast cancers
  - Activating Ras mutations in 25% of melanomas
  - E2A-PBX1 translocation in 5-10% of leukemias



### Biomarker Profile

T-statistic = 12.994

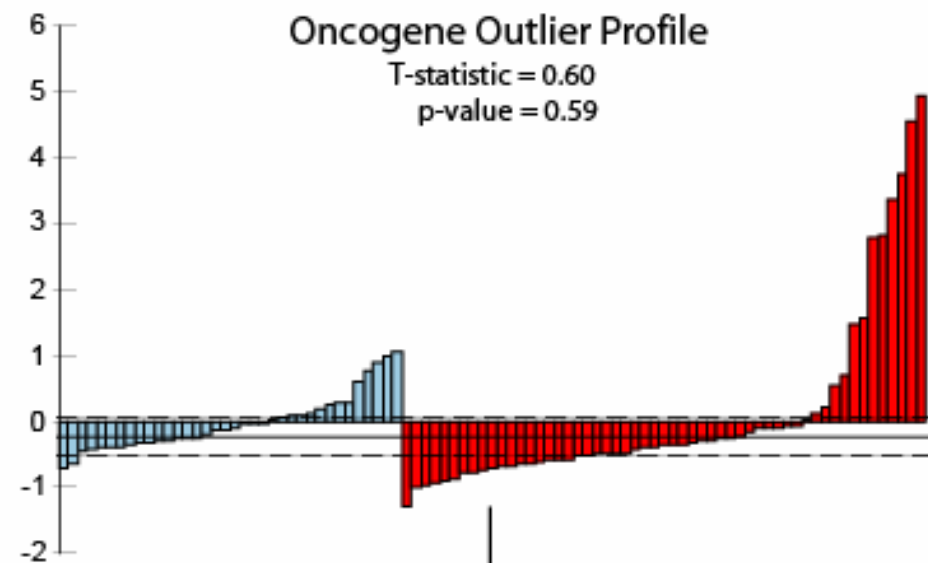
p-value = 1.4E-22



### Oncogene Outlier Profile

T-statistic = 0.60

p-value = 0.59



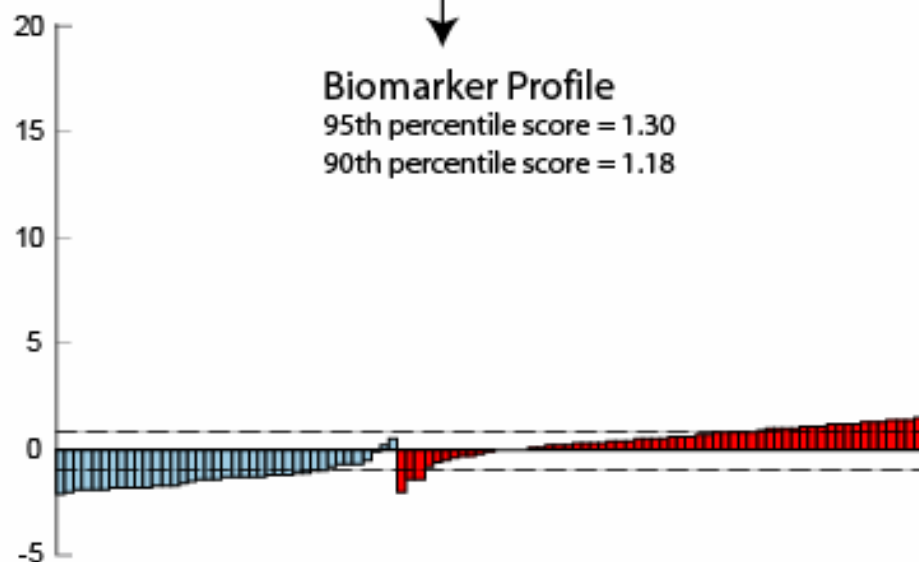
### COPA Transformation

1. center median to zero
2. scale median absolute deviation to one

### Biomarker Profile

95th percentile score = 1.30

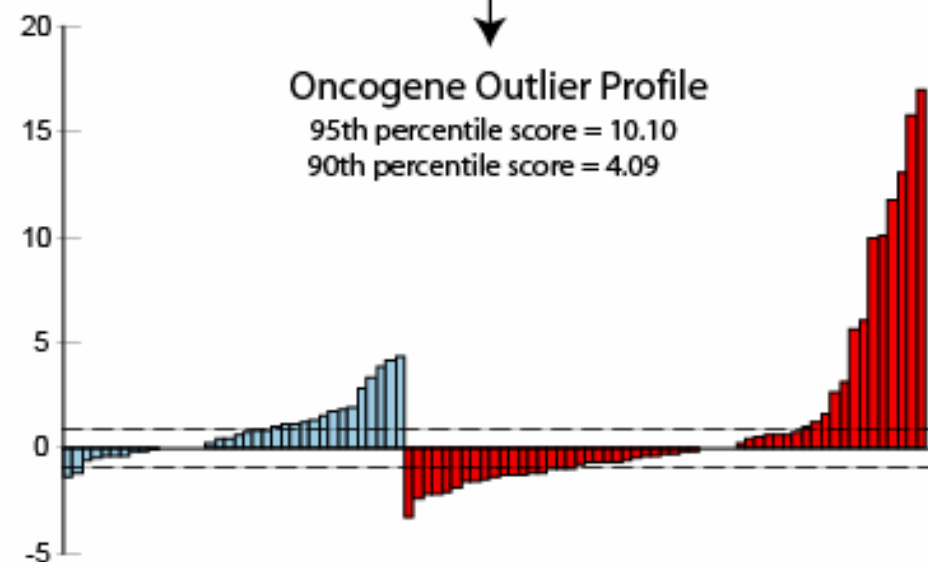
90th percentile score = 1.18



### Oncogene Outlier Profile

95th percentile score = 10.10

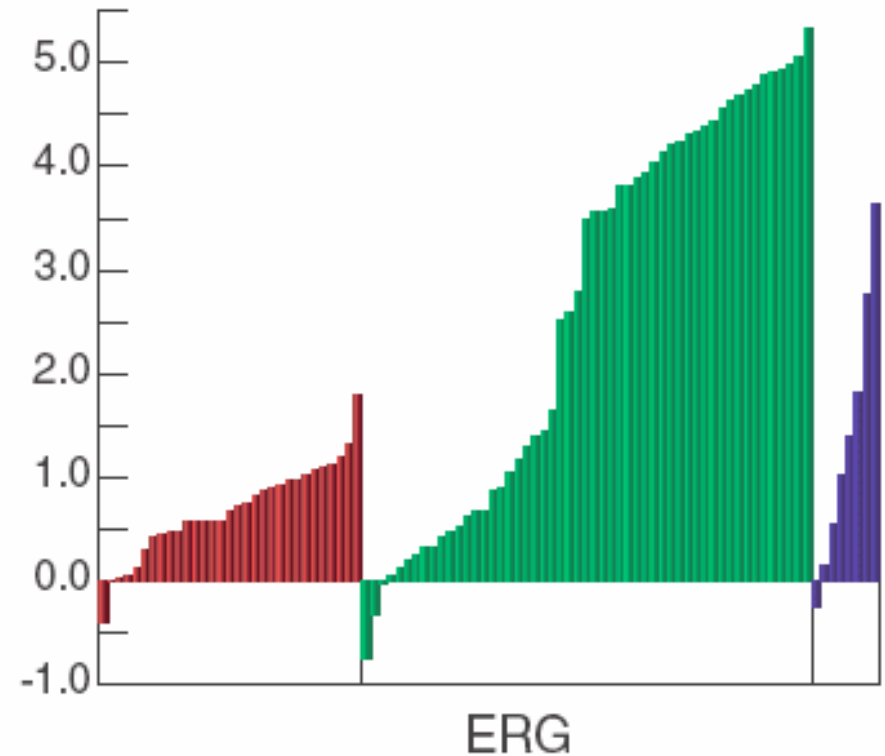
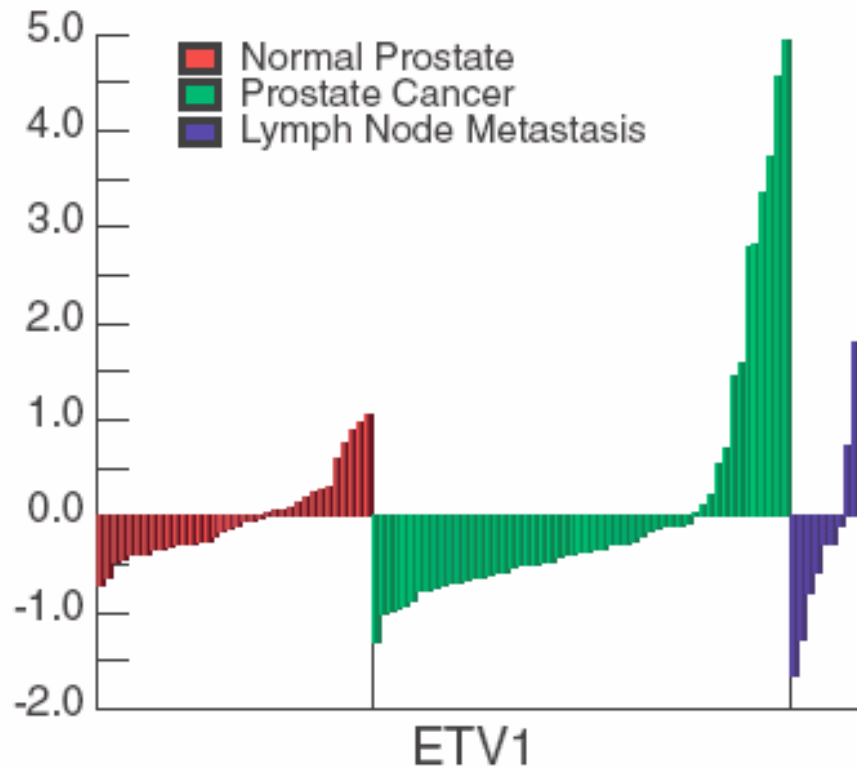
90th percentile score = 4.09



# COPA Results for Several Cancers

Rank	%	Study	Cancer	Gene
1	95	Valk et al.	Leukemia	<i>RUNX1T1</i>
1	95	Vasselli et al.	Renal	<i>PRO1073</i>
1	90	Ross et al.	Leukemia	<i>PBX1</i>
1	95	Lapointe et al.	Prostate	<i>ETV1</i>
1	90	Tomlins et al.	Prostate	<i>ETV1</i>
1	90	Tian et al.	Myeloma	<i>WHSC1</i>
1	75	Dhanasekaran et al. 2	Prostate	<i>ERG</i>
1	75	Welsh et al.	Prostate	<i>ERG</i>
1	75	Zhan et al.	Myeloma	<i>CCND1</i>
1	75	Lapointe et al.	Prostate	<i>ERG</i>
1	75	Dhanasekaran et al.	Prostate	<i>ERG</i>
2	90	Tomlins et al.	Prostate	<i>ERG</i>
3	95	Zhan et al.	Myeloma	<i>FGFR3</i>
4	75	Huang et al.	Breast	<i>ERBB2</i>
6	90	Sotiriou et al.	Breast	<i>ERBB2</i>
9	95	Glinsky et al.	Prostate	<i>ETV1</i>
9	90	Nielsen et al.	Sarcoma	<i>SSX1</i>
9	75	Yu et al.	Prostate	<i>ERG</i>

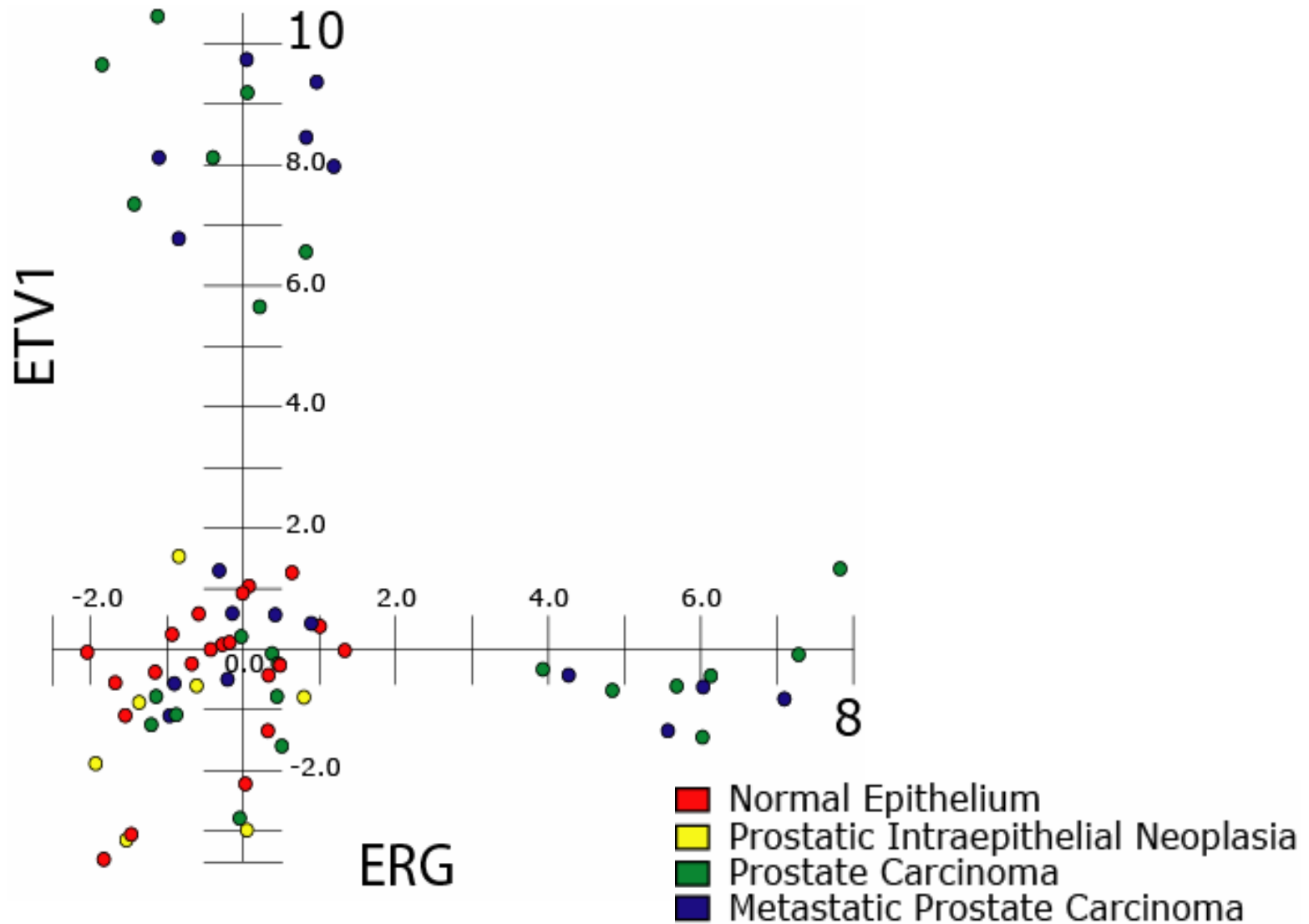
# Outlier Expression of ETV1 and ERG



La Pointe et al

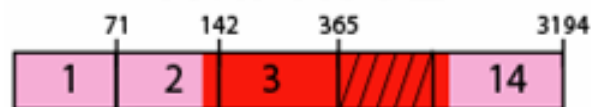
Marked over-expression in a fraction of cases

# Exclusive Outlier Expression of ETV1 and ERG

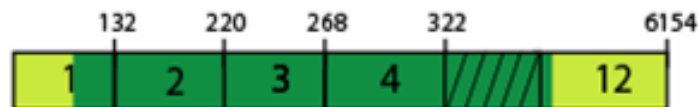




# TMPRSS2



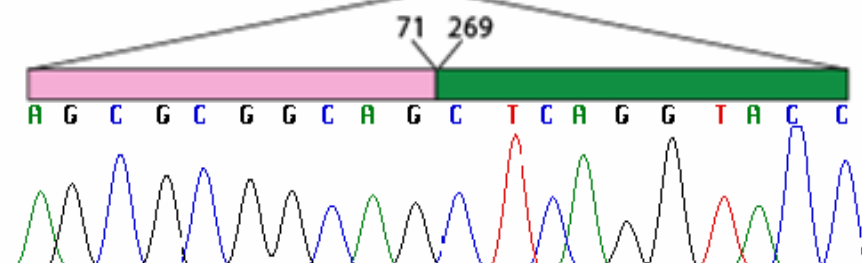
# ETV1



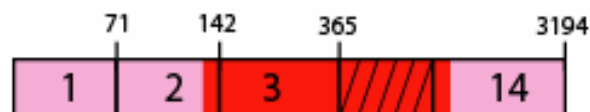
## TMPRSS2:ETV1a



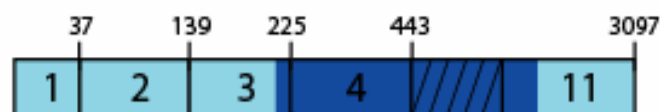
## MET26-LN



# TMPRSS2



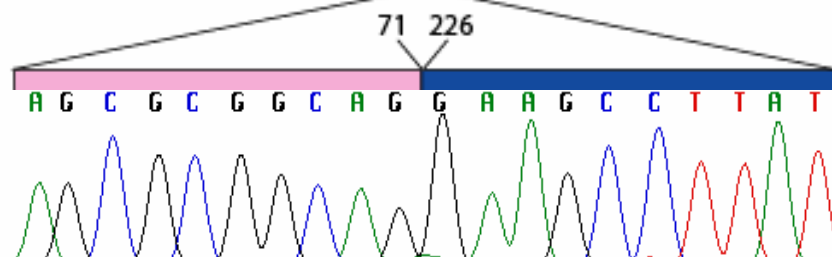
# ERG



## TMPRSS2:ERGa



## MET28-LN



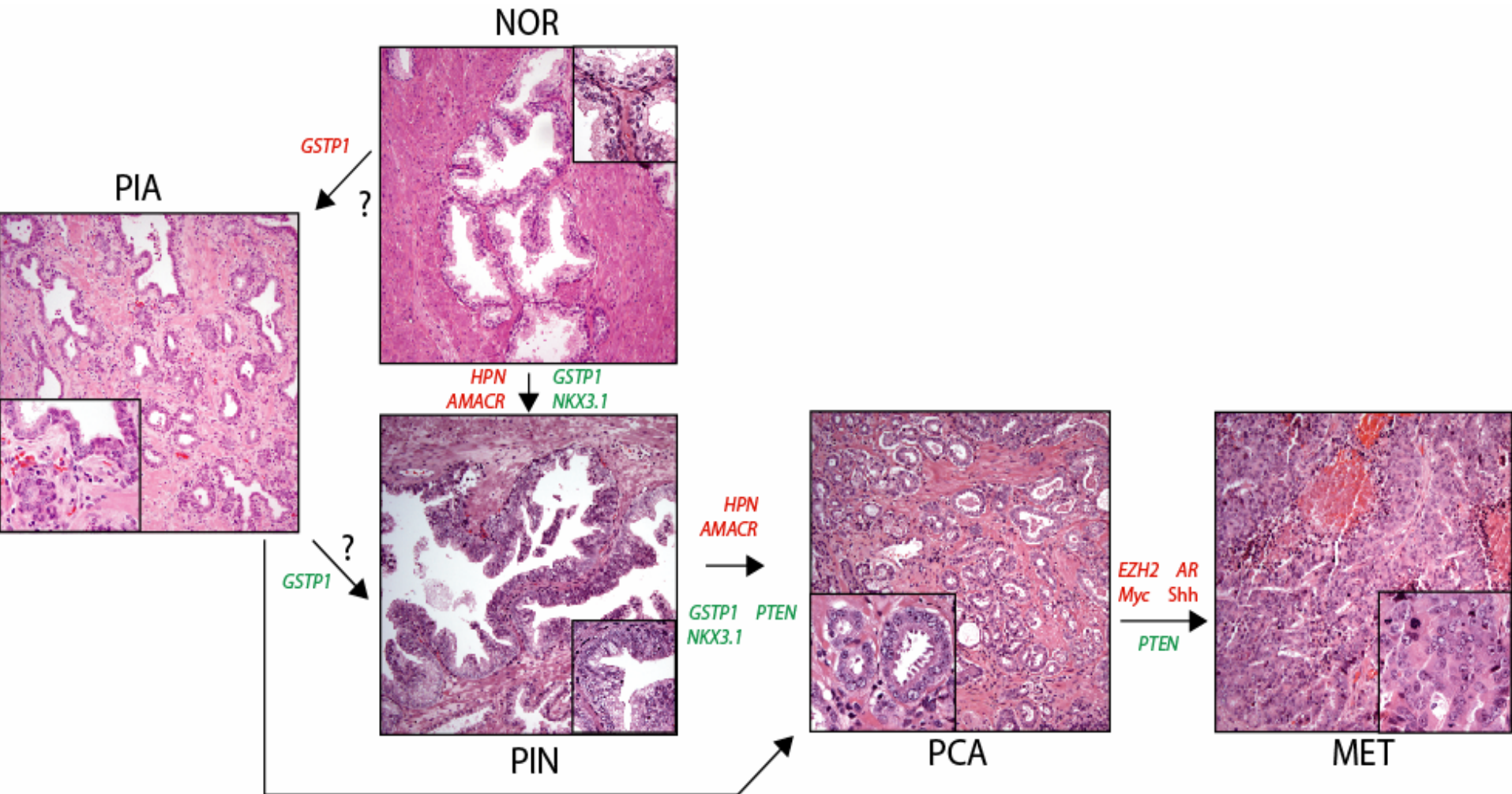
# Conclusions

- Recurrent fusions between the androgen regulated gene TMPRSS2 and ETS family members occur in the majority of prostate carcinomas
- Most common rearrangement in human malignancy

# Informatics Strategies

- Prostate cancer development
  - Informatics strategy to analyze gene expression data
  - (gene fusions)
- Prostate cancer progression
  - Informatics strategy to integrate gene expression data with disparate data sources (molecular concepts)

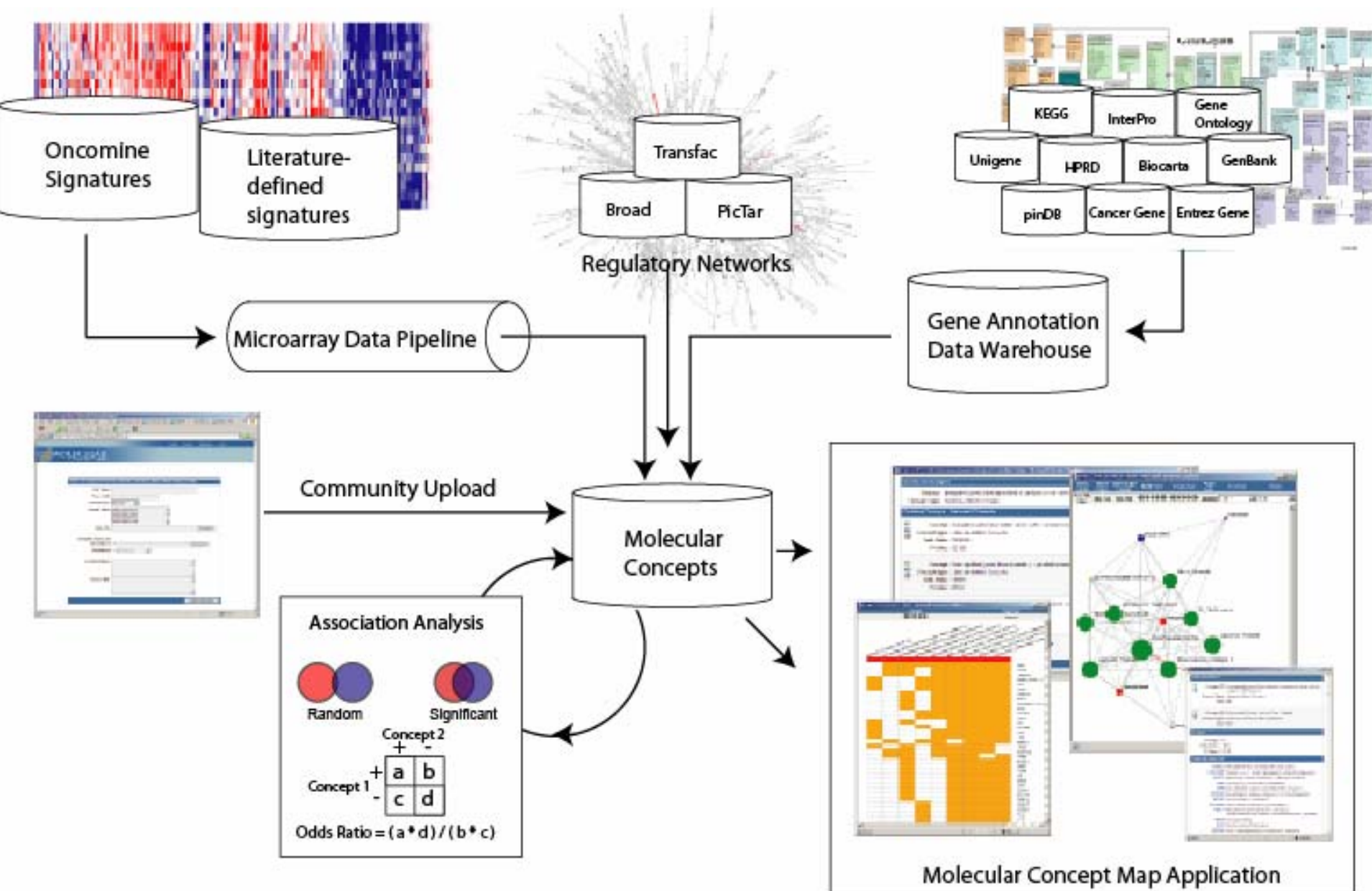
# Prostate Cancer Progression




# Molecular Concept

- A set of genes and proteins related in some biologically-meaningful way
  - Tissue-specific gene expression
  - Shared biological process or pathway
  - Common regulatory element
  - Shared domain
  - Common regulation by an activated pathway or drug treatment

# The Molecular Concept Map Project









[Back](#)[Summary](#)[Filter](#)[Enrichment](#)[Interactome](#)[Pathway](#)[Outlier](#)[PCA](#)**Study:** [tomlinss\\_Prostate\\_Chinnaiyan-Lab](#)**Analysis:** [Prostate](#) - Type2**Class 1:** Normal Epithelium (22)**Class 2:** [Prostatic Intraepithelial Neoplasia](#) (13)**Measured:** 18896**Up:** 1213 (6.4%)   **Down:** 1335 (7.1%) **Diff:** 2526 (13.4%) 

## Profile Summary Report



### GO Biological Process (2 of 1844)


1) 41 of 182 measured genes from [protein biosynthesis](#) are in the top 10.0% of the selected profile. (Q-Value: 1.2E-4)  

2) 16 of 28 measured genes from [proton transport](#) are in the top 20.0% of the selected profile. (Q-Value: 0.003)  

[more.....](#)

### GO Cellular Component (4 of 440)

1) 111 of 365 measured genes from [mitochondrion](#) are in the top 20.0% of the selected profile. (Q-Value: 8.7E-6)  

2) 20 of 83 measured genes from [ribosome](#) are in the top 10.0% of the selected profile. (Q-Value: 0.01)  

3) 13 of 25 measured genes from [proton-transporting two-sector ATPase complex](#) are in the top 20.0% of the selected profile. (Q-Value: 0.001)  

[more.....](#)



### GO Molecular Function (1 of 2086)

1) 28 of 110 measured genes from [structural constituent of ribosome](#) are in the top 10.0% of the selected profile. (Q-Value: 8.4E-4)  

[more.....](#)

### KEGG Pathway (5 of 169)

1) 12 of 55 measured genes from [Ribosome](#) are in the top 5.0% of the selected profile. (Q-Value: 0.008)  

2) 14 of 24 measured genes from [ATP Synthase](#) are in the top 20.0% of the selected profile. (Q-Value: 0.02)  

3) 14 of 24 measured genes from [Photosynthesis](#) are in the top 20.0% of the selected profile. (Q-Value: 0.02)  

**Study:** [tomlinss\\_Prostate\\_Chinnaiyan-Lab](#)

**Analysis:** [Prostate](#) - Type2

**Class 1:** Normal Epithelium (22)

**Class 2:** [Prostatic Intraepithelial Neoplasia](#) (13)

**Measured:** 18896







**Up:** 1213 (6.4%)   

**Down:** 1335 (7.1%) 







**Diff:** 2526 (13.4%) 

## Profile Summary Report

### Literature-defined Concepts (33 of 638)

- 1) 513 of 2045 measured genes from [Upregulated genes in CD4 cells compared to other blood cells](#) are in the top 20.0% of the selected profile. (Q-Value: 3E-12)  
  - 2) 278 of 1020 measured genes from [Downregulated genes in monocytes in response to IL-10 stimulation for 4 to 8 hours](#) are in the top 20.0% of the selected profile. (Q-Value: 4.8E-9)  
  - 3) 65 of 157 measured genes from [Downregulated genes in PC3, DU145, SF359, and U251 cells in response to flavopiridol](#) are in the top 20.0% of the selected profile. (Q-Value: 1.2E-8)  
- more.....*

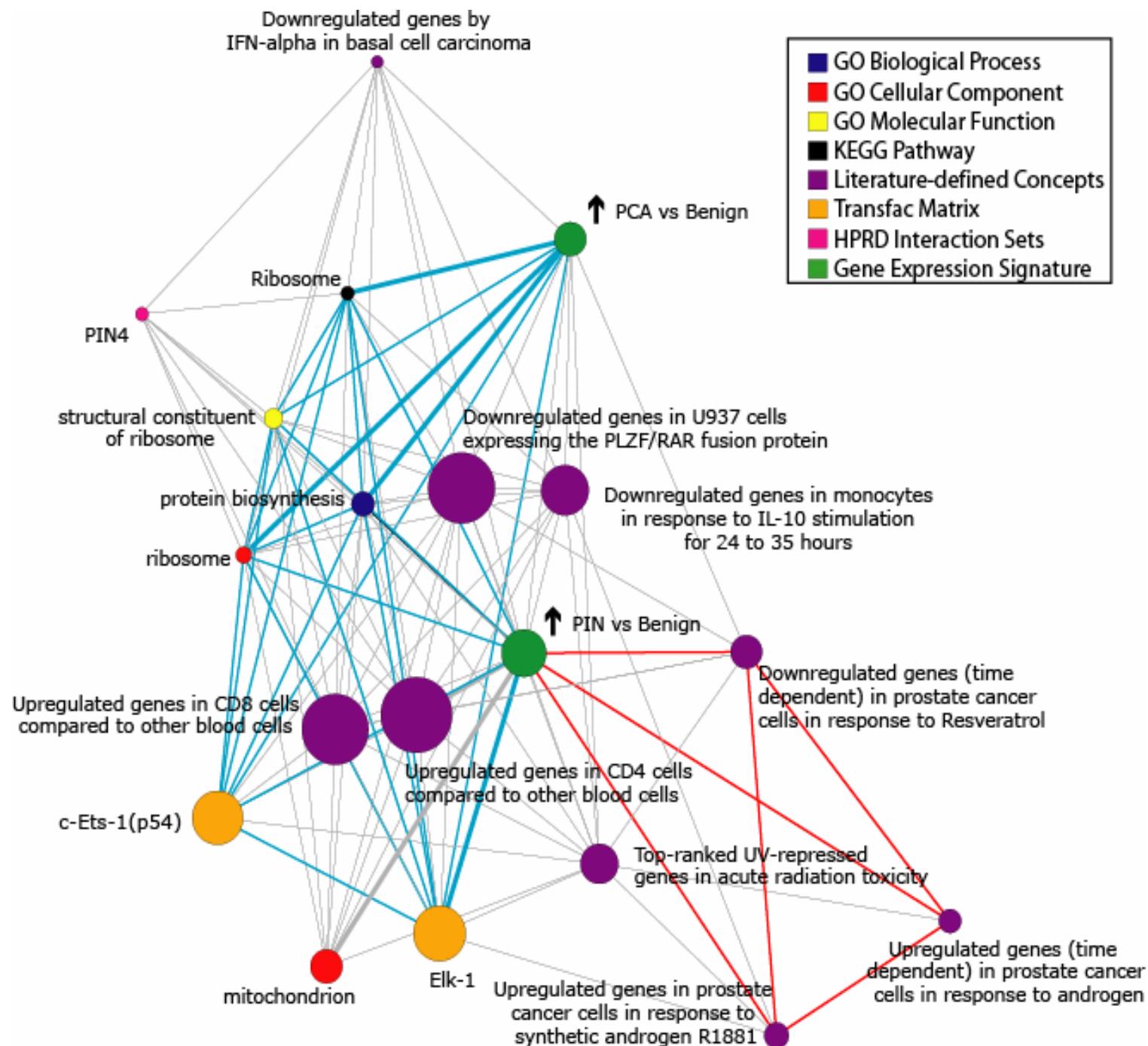
### Transfac TF Matrix - 1000bp (6 of 361)

- 1) 83 of 906 measured genes from [Elk-1](#) are in the top 5.0% of the selected profile. (Q-Value: 7.5E-7)  
  - 2) 130 of 860 measured genes from [c-Ets-1\(p54\)](#) are in the top 10.0% of the selected profile. (Q-Value: 2.4E-6)  
  - 3) 124 of 833 measured genes from [NRF-2](#) are in the top 10.0% of the selected profile. (Q-Value: 9.1E-6)  
- more.....*

### Oncomine Gene Expression Signatures (436 of 5324)

- 1) 364 of 1141 measured genes from [Luo\\_Prostate\\_2::Prostate::Type::Prostate Carcinoma](#) are in the top 20.0% of the selected profile. (Q-Value: 2.7E-33)  
- 2) 185 of 808 measured genes from [Lapointe\\_Prostate::Prostate::Type::Prostate Cancer](#) are in the top 10.0% of the selected profile. (Q-Value: 2.2E-26)  
- 3) 703 of 2952 measured genes from [Bild\\_CellLine::Human Primary Mammary Epithelial Cells::Oncogene Transfected::Control - GSD](#) are in the top 20.0% of the selected profile. (Q-Value: 1.1E-10)  





# Conclusions

- Proliferation of high throughput studies is producing a wealth of data
  - **Analyze and integrate to answer biological questions**
- A novel informatics algorithm was applied across gene expression studies to discover a family of fusion genes as causative events in prostate cancers
- Integrating expression profiling with a compendium of molecular concepts is useful for understanding disease biology

# Next Steps

- How do we interpret the functional significance of the outliers in our profiling study?
  - Know the biology
- Transcriptome, regulatory networks and proteome based approaches
- There is no one right way to analyze and integrate data
  - Methods are complementary

# Acknowledgements

## Arul Chinnaiyan Lab

### **Daniel Rhodes**

Soory Varambally  
Saravana Dhanasekaran  
Terrence Barrette  
Qi Cao  
Arun Sreekumar  
Bharathi Laxman  
Vasu Mahavisno  
Julie Kim  
Rong Zhou  
Lucy Wang  
K. Shanker  
Adai Vellaichamy  
T. Rajendiran  
Tim Bradford

Rohit Mehra  
Xuhong Cao  
Anjana Menon  
Beth Helgeson  
Jianjun Yu  
Radhika Lnu  
Ronglai Shen  
George Wang  
Chad Creighton  
Jindan Yu  
Dave Hanauer  
Barry Taylor  
Rui Wang  
Jane Cao

## Michigan

Rajal Shah  
Ken Pienta  
Eric Fearon  
John Wei  
Jim Montie

## Harvard

Mark Rubin

## Support

**NCIBI**  
**SPORE**  
**EDRN**  
**MSTP**



### User Profile

**tomlinss@med.umich.edu**

[Change password](#)

### New Studies

[Bild\\_CellLine](#) 

[Ma\\_Breast\\_2](#) 

[Dong\\_Brain](#) 

[Browse All](#)

### Statistics

Studies - Catalog: 896

Studies - Analyzed: 149

Microarrays: 16656

Data points: 308226536

Cancer Types: 49

Registered Users: 8416

### Highlights

[Mining for Regulatory Programs](#)

Oncomine recognized in  
[IAMA](#)

## GENE SEARCH



(e.g "CDK4","ERBB2","IMAGE:417226")



## PROFILE SEARCH



(e.g "prostate","grade","glioblastoma")



### About Oncomine

Welcome to Oncomine (3.0), a resource for examining gene expression in cancer. The goal of the project is to collect, standardize, analyze, and deliver published cancer gene expression data to the research community. Probe the expression of a gene across thousands of cancer samples or explore genes, processes, and pathways deregulated in a particular type of cancer. Oncomine pre-computes cancer profiles, clusters, and gene set modules so you can focus on discovery. Read more [here](#).

Oncomine was developed by physicians, scientists, and software engineers at the University of Michigan. Direct questions and comments to [oncomine@umich.edu](mailto:oncomine@umich.edu)

**Commercial users**, please visit [www.oncomine.com](http://www.oncomine.com) for licensing details

Gene Symbol: ERBB2 

Gene Name: V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)

## Differential Activity Map



P-Value Threshold: 1E-4

+

-

Outlier Rank Threshold: 50

+

-

	Normal vs. Normal	Cancer vs. Normal	Cancer vs. Cancer	Molecular Alteration	Misc	Outlier
Kidney	2					
Vulvar	1					
Brain	1	1	2		2	
Lung		2	2 3			
Ovarian		1				
Lymphoma			1			2
Prostate			2 1		1	
Renal			1 1			
Sarcoma			2		1	
Breast				1 1		6

**Gene Symbol:** ERBB2 

**Gene Name:** V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)

Filter By



Results 1 - 20 of 22 [All > Breast >](#)

[Next 2 >>](#)

**Study:** [Huang\\_Breast](#)   

**Reporter:** [33218\\_at](#)

**Tissue:** [Breast Carcinoma](#) (89)

**Value:** #4 outlier at the 75th percentile (score = 2.5728)

**Study:** [Sotiriou\\_Breast\\_2](#)   

**Reporter:** [IMAGE:783729](#)

**Tissue:** [Breast Ductal Carcinoma](#) (99)

**Value:** #6 outlier at the 90th percentile (score = 6.6079)

**Study:** [Huang\\_Breast](#)   

**Reporter:** [1802\\_s\\_at](#)

**Tissue:** [Breast Carcinoma](#) (89)

**Value:** #14 outlier at the 75th percentile (score = 2.2698)

**Study:** [Sotiriou\\_Breast\\_2](#)   

**Reporter:** [IMAGE:141768](#)

**Tissue:** [Breast Ductal Carcinoma](#) (99)

**Value:** #15 outlier at the 90th percentile (score = 5.4415)



## Sotiriou\_Breast\_2

( Tissue: Breast Ductal Carcinoma, Parameter: Grade)

## ERBB2

V-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homol

